Methods in epidemiology and public health: does practice match theory?

D L Weed

Epidemiology is widely perceived as a public health discipline within which methodology matters. Methods dominate educational curriculums and influential textbooks. Epideriometological societies regularly feature methods sessions at their national and international meetings and, at least informally, the discipline recognises the methodologists who study the methods and the practitioners who use them. It follows that epidemiological methods, whether quantitative (for example, meta-analysis or logistic regression) or qualitative (for example, causal inference or narrative reviews), have a theoretical side and a more practical side. The theory behind a method and especially how that method should be practised “in theory” are discussed in textbooks and journal articles; methodological standards and guidelines are good examples. How a method is actually practised is found in its applications, in analytical studies, reviews, and other publications. The relation between these accounts—the extent to which practice matches theory—is the starting point of this paper.

Three methods are of particular interest: meta-analysis, causal inference, and techniques for systematic narrative literature reviews. These methods were selected because serious concerns have been raised about how each is practised. These are also closely related methods, often appearing together in the same publication, typically a review paper, textbook chapter, or technical report within which a body of evidence is summarised and interpreted. Nevertheless, each of these methods is distinct enough to have a recognisable theoretical literature and observable practice patterns.

Perhaps the most important reason for looking carefully at the relation between these methods is how they are “supposed” to be practised is the key part they play in the assessment and interpretation of scientific evidence. These methods are central to the search for causal determinants of disease and for ways to use that knowledge to improve public health.

Public health and clinical disciplines are joined by education, psychology and the social sciences in using these methods. Practitioners from many disciplines may therefore gain insight from this methodological inquiry, that begins at the juncture of theory and practice (do they match?) and uses contemporary philosophical concerns to illuminate what is uncovered there and to help lay out a course for the future.

Do theory and practice match?

Inquiries into the relation between any method as a method and the practice of that method are relatively uncommon. There are, however, a few recent examples in epidemiology. Very little mismatch was observed between the methodological accounts of case-control studies of cancer screening and their practice, perhaps because those who write the methods papers in this highly specialised area are also directly involved in designing investigations of screening programmes using the case-control approach. For other methods, the situation is not so neat. Considerable mismatch between methodological standards and actual practice was recently identified in clinical epidemiological studies of molecular and genetic factors.

For the three more interpretative methods to be examined in this paper, inquiries into the relation between practice and theory reveal a complex and intriguing array of issues. I begin with meta-analysis.

META-ANALYSIS

Meta-analysis, especially for clinical trials, has a growing number of textbooks and journal articles that outline a generally acceptable approach for applying it. And while this approach may not be precisely the same across different accounts (even correcting for the dynamic and progressive nature of scientific thinking) it is reasonable to assume that in any relatively narrow time frame the methodology of meta-analysis applied to clinical trial data will probably differ in minor rather than fundamental ways. Nevertheless, Bailar has shown in his informative if unsystematic review of the practice of meta-analysis that vivid examples of inadequate even careless applications of the technique exist, clearly in contrast with what a reasonable reader of the methodological literature would conclude represents “acceptable” practice. A study of the practice of meta-analysis in epidemiology supported this finding, revealing that only about one third of published examples assess heterogeneity, a central concern of the methodological literature.

CAUSAL INFERENCE METHODS

For causal inference methods, the practice is also variable, but then so are the methodological accounts. Although Hill’s classic list of criteria often represents the core set of concepts for discussions and applications of this qualitative method, the authors of textbooks and methodological articles often select different subsets of Hill’s list and assign different rules of evidence to them. The practice of causal inference reflects this value laden and subjective selectivity not only in how the criteria are
chosen and defined, but also in how practitioners evaluate bias, confounding, and the relative evidentiary importance of different study designs.15 16

SYSTEMATIC NARRATIVE REVIEWS
So-called systematic narrative literature reviews are practised with no more methodological consistency than are meta-analysis or causal inference. In a recent study of the quality of all review papers published in 1995 in all major epidemiological journals, 60% had major flaws. Lack of a stated purpose, inadequate descriptions of literature search techniques, and failure to disclose the studies included and excluded topped the list of deficits.6 For narrative reviews, rather than highly variable methodological accounts being highly variable practice as described for causal inference, there seems to be a lack of attention by authors, editors, and peer reviewers to methodological standards proposed and discussed for at least a decade.17 18

Mismatches reconsidered
Explanations of these phenomena are probably more worth more space than is allotted here, but a few ideas are worth considering. For the case of systematic reviews, the presence of methodological standards seems not to have influenced the authors’ largely narrative approach. A contributing factor may be that reviews have long been considered second class citizens to primary research articles, more personal reflection than scientific evaluation.19 It follows that practitioners may not feel much need to systematise their approach, even if reasonable guidelines exist. Such a sentiment, and indeed the practice itself, would probably change if journals required systematic approaches to narrative reviews as a condition for publication.

The situation for meta-analysis, especially when applied to observational data, is complicated by the fact that the accounts describing how (even whether) meta-analysis should be performed are so diverse that in a single journal issue it is easy to find a paper cautiously extolling the virtues of the technique alongside another claiming that meta-analysis of observational data is equivalent to statistical hocus pocus practised by investigators who should know better.20–22 Feinstein, for example, calls meta-analysis “statistical alchemy”.23 Small wonder the practice of meta-analysis is so varied, even completely ignored, by some who review observational evidence. Standards for the practice of meta-analysis24 may help, assuming that practitioners follow them. Meta-analysis is, after all, primarily a statistical technique and although not completely driven by algorithms and highly specified quantitative rules, it is nevertheless more amenable to standard setting than are the qualitative methods of causal inference. Meta-analysis need not become a fixed routine, but rather its practitioners could strive to become, in Bailar’s words, “adequately qualified” to perform it (page 155).4

It seems reasonable to conclude for the methods of narrative reviews and meta-analysis that existing guidelines and standards should be better followed by those who practise these methods. This recommendation does not discount the important part that judgment and other more philosophical concerns play in the application of methods, whether quantitative or qualitative. To these I will return.

A deeper problem for causal inference
The problem of variability in the practice of causal inference is more troubling than that found in systematic reviews and meta-analysis primarily because the methodological accounts of causal inference mimic and thereby support a subjective—highly variable—practice. Guidelines are not currently available that could provide a nuanced and comprehensive approach to causal inference, including standards for its many specific components (for example, causal criteria).1 Quantitative alternatives may sound appealing, but again no useful proposals have been offered, reflecting the difficulties of the task, which derive from the fact that the basic questions involved in all such exposure-disease assessments, questions such as, “is the exposure causal?” and “should we recommend public health action?” are not quantitative. Certainly there are quantitative issues involved, such as the magnitude of the relative risk for causation and the magnitude of the absolute risk for public health recommendations, but the admixture of qualitative concepts and quantitative measures and the infusion of both scientific and ethical reasoning in the process suggests that an appeal to methodological standards is not going to work without first considering the philosophical groundworks upon which such standards could be constructed.

Philosophical considerations: the current landscape
Put another way, we can examine the methodological problem described above from a deeper—that is, philosophical, perspective. Philosophical issues and concerns may provide insight and understanding.28 Philosophy is, after all, comprised of more fundamental constructs and concepts upon which the methodologies and practices of a public health discipline rely. Importantly, causal inference methods have already been linked to some philosophical constructs, in particular, epistemological models and ethical decision making approaches.

The reader should keep in mind, nevertheless, that philosophical inquiry has its share of blind alleys and false hopes. My purpose here is primarily to briefly examine the extent to which philosophy may assist us in better understanding why the current practice and theory of causal inference is rife with variability, suggesting what we may need to do—and by “we” I mean anyone who uses these methods—to make progress, and thus paving the way for others to develop these ideas more fully. By taking a philosophical approach to understand the problem of variability in causal inference methods, our understanding of the mismatch phenomena for the other
methods discussed above may also be improved, and suggestions for improvements in theoretical and practical levels. Although we can only glimpse a few notable landmarks, clearly the concepts of causality and prevention lie at the heart of it. Epidemiologists purport to study the determinants of disease and apply that knowledge to improve public health through preventive interventions. Causal inference methods provide solutions to the paired problems of making causal conclusions and public health (preventive) recommendations from scientific (epidemiological, biological, clinical, and social) evidence. Causal theory, in turn, is a rich philosophical arena characterised by many—deterministic, probabilistic, and counterfactual—approaches. Indeed, if the epidemiologists' views on causality vary across these categories, such that different practitioners have different causal hypotheses in mind when they interpret the evidence, then it might be reasonable to expect a concomitant variability in the practice of causal inference. Unfortunately, there are few published connections between causal theories and causal inference methods and none rooted in actual practice. If variability in epidemiologists' allegiances to causal theory is linked to variability in their practice of causal inference, it is a matter of conjecture.

Leaving the realm of ontology for epistemology and ethics—these three being the domain of any philosophical approach to a public health discipline—the roots of the problem of mismatch become clearer. The main reason why variability in the practice of causal inference is so troubling is that we may reasonably expect from these methods causal judgments that are correct (in the case of scientific conclusions) and ethically justified (in the case of public health or preventive recommendations). Simply put, it matters a lot that from a review of epidemiological evidence conclusions about causation and public health recommendations arise and it matters even more that highly variable methodological practices beget highly variable conclusions. Examples of such variability—extraordinary differences in causal conclusions and public health recommendations emerging from different investigators examining the same evidence—include controversial reviews published on abortion and breast cancer, vasectomy and prostate cancer, smoking and cervical cancer, alcohol and breast cancer, and the use of mammography for women under 50 years of age. While these are only a few examples from this author's experience, the implications for public health and for medicine at large are real and troubling. The implications for the public and its primary source of epidemiological information, the media, are even more daunting. The public and media's perceptions of epidemiological science and its role in public health are confused and perhaps a bit hostile.

Epidemiologists and the media and public and indeed, other scientists, want to know if causal claims are "real"; they also want to know if these same claims, when translated into public health recommendations are just and beneficial.

From a philosophical perspective, the current situation can be distilled into two candidate issues: (1) whether judgments about causation can be held up against an objective evaluative standard and (2) whether public health and medical practice recommendations can be justified by some secular (that is, non-sacred) bioethical framework. Unfortunately, those who think hard about the existence of such entities—the philosophers of science and bioethicists—are not convinced that either exists. For science and its philosophical foundation, a truly objective standard (the "real" in realism) is only one possibility against which the neopragmatists align their consensus-based approaches featuring what "works" rather than what "is". Postmodernists, in turn, posit their historically sensitive and socially mediated paradigms that rise and fall like century long intellectual tides, providing no firm foundation if new paradigms are incommensurable with the old. If any reader doubts that these sorts of concerns affect epidemiological thinking, see the recent series of papers on paradigms and pragmatism in epidemiology.

In bioethics, some thinkers argue that we are awash in diverse moral perspectives—it is a radically pluralistic jungle out there—afflicting all levels of moral discourse from the application of moral reasoning in particular cases to the very foundation of theoretical and applied bioethics. These problems are too dense for unpacking here but cannot be ignored or dismissed as irrelevant "philosophical" musings. Increasingly, it is recognised that implicit philosophical commitments often affect practical decision making.

So here is the rub as we arrive full circle from where we began this philosophical excursion: if the foundations of scientific and public health decision making methods have no firm consensual (nor rational) basis in philosophy itself, and if the methods themselves in turn support a subjective practice, then we should certainly not expect the decisions that arise from those methods to be consistent, much less be "objective" or "ethically justified."

"High, low, and middle ground" philosophy

Before we can move forward from this rather bleak and precarious intellectual position, we may want to reconsider carefully how philosophy can help us. We are primarily concerned about the connection between the practice of a method and the more theoretical accounts purporting to describe and prescribe that same practice. Blackburn provides an interesting and relevant metaphor; he writes that philosophy can be done on the "high, middle, or low" ground. The first of these is what real philosophers do and have done for millennia when they discuss issues at levels well above anyone's
Methods-practice mismatch

Virtues and the problem of mismatch
I begin with MacIntyre's decade old thesis that at least one philosophical theory can claim a serious longevity, remaining as it does not so much hidden as camouflaged beneath centuries of neglect and intellectual trends. Virtue theory, with its emphasis on character and motivation, and its relevance to both science and ethics may be the raw material from which scientists, public health practitioners, and therefore epidemiologists can rebuild their philosophical arks and so their methodological fortresses. That valuable and complex effort is well beyond the scope of this paper, but it will emphasise traits such as prudence and benevolence, integrity and excellence.

One leg in our philosophical journey, therefore, must involve examining ourselves at the level of the individual professional, encouraging those we mentor and train through teaching and by example what it means to be a good epidemiologist, committed to prevention and a benevolent and just application of scientific knowledge.

Casuistry (case-based moral reasoning) and the use of causal inference methods
Another philosophical avenue worth exploring is found in contemporary bioethics. Decisions about public health recommendations in the process of causal inference may be made, as are many ethical decisions, using another long-standing philosophical technique, that of case-based reasoning or casuistry. The method of casuistry emphasises the importance of the particular circumstances of decision making, with various ethical rules or even principles selected for application depending upon the situation. This is precisely the sort of decision making that takes place in contemporary epidemiology when causal inference methods are invoked, with different causal criteria and different rules of inference being selected and prioritised depending upon the particular circumstances of the disease-exposure association under consideration. That the casuist technique has been connected to MacIntyre's virtue-based theory along communitarian lines and that it is complementary to the principle-based approach to ethics that epide-
Improve the connection between theory and practice, we may wonder if progress, nevertheless, is being made. Clearly, the (tentative) solution of virtue ethics brings up a new problem: how to teach virtues, in as much as they are traits of character rather than skills or cognitive processes. A (tentative) solution to that problem, even at a the highly specific level of a particular individual scientist, involves the presence of a mentor of high moral fibre who, by example, can provide a model of excellence for her student. Even recognising the possibility of slipping along a path towards regress, I believe that progress in understanding the problem of mismatch is being made and that Popper’s scheme captures its essence in a way that is useful. For an account of this same problem solving scheme applied to causal criteria, see Weed. For a metaphorical account, see Weed. For an account of applying a strict Popperian refutationist view to meta-analysis, see Maclure.

Conclusions

Mismatches between methods and practice for three closely related interpretative methods in epidemiology occur for a variety of reasons and may therefore respond to several different strategies. For causal inference, a lack of objective methodological standards and the spell of postmodernism contributing to a permissive interindividual value laden subjectivity—a “do as you please approach”—affecting authors, peer reviewers, and editors alike explains some of this variability. For the qualitative method of systematising literature reviews, it seems simpler to describe and harder to justify: we have been slow in applying reasonable and defensible methodological rules. Is it a failure of communication, for example from editors to authors, that prevents us from improving reviews? Or is it that we suffer from some pervasive motivational malady of professional character; we know we should systematise our reviews, but we just cannot bring ourselves to do it? As for meta-analysis, it seems that the controversy surrounding the technique, especially among epidemiologists, along with inattention to existing standards, has contributed to a practice environment with too much room for individual preferences.

So where do we go from here? An obvious direction is to do a better job in educational and training programmes, providing to students and trainees the existence of and need for methodological guidelines as well as a better understanding of the nature and use of professional judgment. But this approach begs the question unless more specific recommendations ensue and unless we can better characterise the slippery notion of judgment itself. My view is that, at least for the present, the methodologies of meta-analysis and systematic reviews have made some progress and that we should better use (and therefore teach) existing methodological standards. In short, let us practise what we preach. But the same cannot be said for causal inference methods. We much preach and practise a poorly supported and barely reasoned methodology. Progress towards standards for causal inference methods...
seems more challenging and potentially more rewarding; it is no simple task to lay out standards for causal claims and public health decisions, dependent as they are upon our fundamental and often unstated philosophical perspectives, implicit causal models, historical experience, paradigmatic cases, and judgment. Philosophical inquiry will be important. In this paper, virtue theory, casuistry, and a problem solving technique within critical rationalism were used to provide insight and guidance into methodological concerns. There are many other choices.

Although this paper has focused upon interpretative methods in epidemiology, the relation between theory and practice for other methods is worthy of attention. How studies are designed and statistics calculated, how interactions are assessed and p values calculated, and how logistic regression and structural equations are applied, can all be considered both how logistic regression and structural equations are assessed and p values calculated, and how interactions are assessed and p values calculated. There are many rewarding; it is no simple task to lay out standards for causal claims and public health decisions, dependent as they are upon our fundamental and often unstated philosophical perspectives, implicit causal models, historical experience, paradigmatic cases, and judgment. Philosophical inquiry will be important. In this paper, virtue theory, casuistry, and a problem solving technique within critical rationalism were used to provide insight and guidance into methodological concerns. There are many other choices.

Although this paper has focused upon interpretative methods in epidemiology, the relation between theory and practice for other methods is worthy of attention. How studies are designed and statistics calculated, how interactions are assessed and p values calculated, and how logistic regression and structural equations are applied, can all be considered both how logistic regression and structural equations are assessed and p values calculated, and how interactions are assessed and p values calculated. There are many rewarding; it is no simple task to lay out standards for causal claims and public health decisions, dependent as they are upon our fundamental and often unstated philosophical perspectives, implicit causal models, historical experience, paradigmatic cases, and judgment. Philosophical inquiry will be important. In this paper, virtue theory, casuistry, and a problem solving technique within critical rationalism were used to provide insight and guidance into methodological concerns. There are many other choices.

Although this paper has focused upon interpretative methods in epidemiology, the relation between theory and practice for other methods is worthy of attention. How studies are designed and statistics calculated, how interactions are assessed and p values calculated, and how logistic regression and structural equations are applied, can all be considered both how logistic regression and structural equations are assessed and p values calculated, and how interactions are assessed and p values calculated. There are many rewarding; it is no simple task to lay out standards for causal claims and public health decisions, dependent as they are upon our fundamental and often unstated philosophical perspectives, implicit causal models, historical experience, paradigmatic cases, and judgment. Philosophical inquiry will be important. In this paper, virtue theory, casuistry, and a problem solving technique within critical rationalism were used to provide insight and guidance into methodological concerns. There are many other choices.
60 Weed DL. Epidemiology, the humanities, and public health. *Am J Public Health* 1995;85:914–18.